SHRP 2 Reliability Project L13A

System High-Level Architecture

SHRP 2
STRATEGIC HIGHWAY RESEARCH PROGRAM
Accelerating solutions for highway safety, renewal, reliability, and capacity

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

SHRP 2 Reliability Project L13A

System High-Level Architecture

Prepared by
Iteris

TRANSPORTATION RESEARCH BOARD
Washington, D.C.
2015
www.TRB.org

ACKNOWLEDGMENT

This work was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials. It was conducted in the second Strategic Highway Research Program, which is administered by the Transportation Research Board of the National Academies.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. (Dan) Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Victor J. Dzau is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C.D. (Dan) Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

Table of Contents

1.1 INTRODUCTION

This document outlines all the components of the SHRP 2 Archive System. Most of the content included in this document is a part of Chapter 7 of the final L13A report. System and other technical details provided in Chapter 7 of the final report were reviewed and revised in this document to reflect the most up-to-date details.

SHRP 2 Archive System consists of the following components:

- Amazon Web Services (AWS);
- Apache HTTP server;
- WordPress system with specific SHRP 2 plugins and themes;
- MySQL database;
- Tomcat application server;
- Solr search engine;
- S2A Server; and
- Tripwire.

These components are interconnected as shown in Figure 1-1. The data flow between these components is depicted in Figure 1-2.

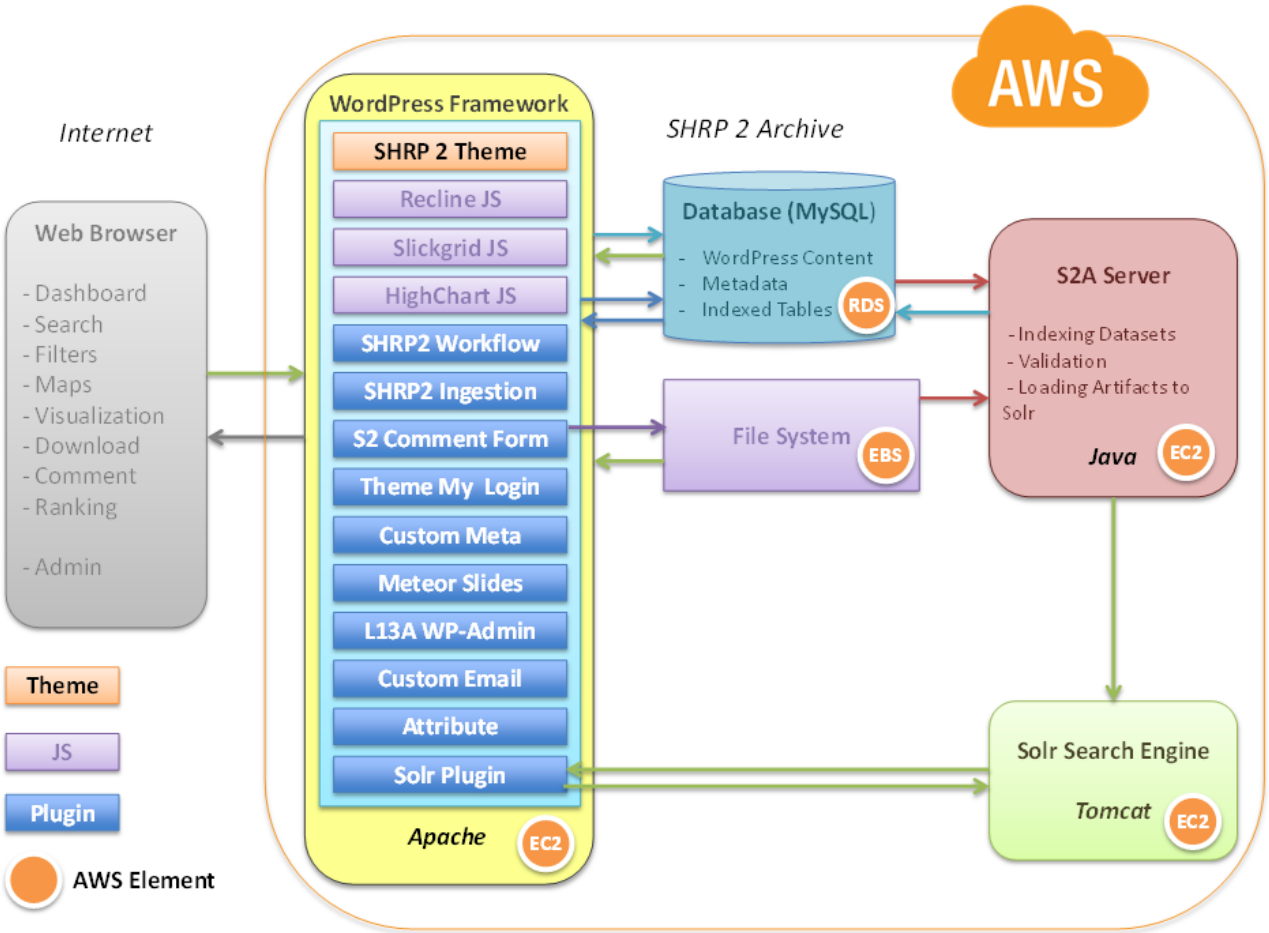

Figure 1-1: Components of SHRP 2 Archive.

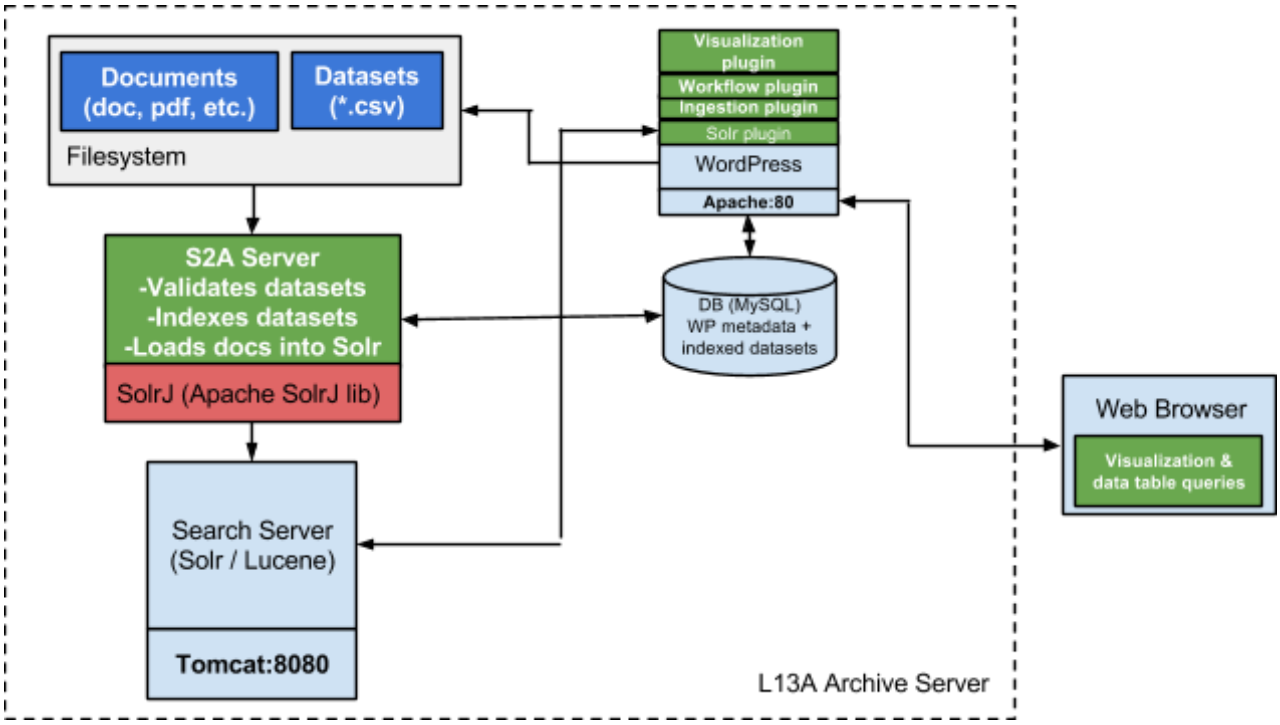

Figure 1-2: Data Flow Diagram for the L13A Data Archive.

The following sections provide detailed information on each of the above listed components.

1.2 AMAZON WEB SERVICES (AWS)

AWS is a bundle of remote computing services that provides a cloud-computing platform that is offered over the Internet. Both the L13 report and L13A Iteris team's assessments indicated that the cloud-based service is a viable solution for hosting the Archive. From the Iteris team's point of view, the proposed L13 architecture (See Section 3.1.10 of the L13A report) was slightly outdated. To that end, the Iteris team modified the proposed architecture and leveraged the extensive cloud-based services Amazon provides to the public. We deployed the Archive system on a bundle consisting of the following components:

- **Amazon Elastic Compute Cloud (EC2)**. EC2 provides virtual servers and is delivered on the CentOS (RHEL) operating system. EC2 manages the data and information via Elastic Block Storage (EBS). EBS is a volume-based storage that has a separate life span and can be attached to any instance. EBS module size is 200 GB and can be resized. For now the team has used the medium M3 instance for the EC2 module. It should be noted that in our design we have not implemented a hot standby instance as a backup for cases when the operation of the EC2 module fails. Amazon guarantees uptime of more than 99%. In case of any potential failure the admin team can set up another instance in a couple of hours.
- **Amazon Relational Database Service (RDS)**. Database administration (e.g., configuration, backup, monitoring resource consumption, etc.) is an expensive and error-prone task. The purpose of this module is to provide a relational database service via amazon cloud that helps users save money and avoid errors. RDS supports three popular relational databases, i.e., MySQL, SQL Server, and Oracle. The Archive utilizes MySQL for managing its database

system. As of September 2014, the size of the database was 500GB. The service is elastic. Therefore scaling up the storage is very easy.

1.3 WORDPRESS

WordPress is one of the most popular open source content management and blogging systems available. WordPress was selected as the core Content Management System (CMS) of the Archive after a thorough assessment of various Commercial Off-the-Shelf (COTS) CMSs (see Section 3.5.4 of the L13A report for more information).

WordPress requires a web server with PHP support, a URL rewriting facility, and an instance of MySQL. The Archive system uses Apache as the HTTP server. Apache is a preferred option developers normally implement with WordPress because it provides PHP interpretation and URL rewriting.

1.3.1 Themes

The WordPress theme is the face and graphical aspect of the website which encompasses the entire user experience. Therefore, the appearance of the user interface is built based on a theme. A theme is a bundle of template files (PHP files to provide logic and structure), CSS files (to keep the style), images, and JavaScripts.

There are many WordPress theme resources available that can be used directly or customized. The SHRP 2 Archive theme is a child theme of WordPress' Twentyeleven general theme. The SHRP 2 Archive theme was customized for the Archive user interface.

1.3.1.1 Key Open Source JavaScript Libraries Used in the Archive

JavaScript works within WordPress. It can be used within WordPress template files in WordPress Themes or Child Themes. As recommended by the L13 report, the Iteris team aimed to use open source libraries as much as possible. Table 1-1 summarizes the list of open source JavaScript libraries used to deliver some of the core functionalities of the Archive system.

Table 1-1: List of open source JavaScript Libraries

JavaScript	Description
Recline	Library to build data applications. It can be integrated with Leaflet, Slickgrid, and Highcharts. This library was used a platform that delivers the visualization functionalities on the Data tab located on top of the data set pages.
Slickgrid	Grid/spreadsheet view of the data sets.
Highchart	Data set plots and graphs.
Leaflet	Interactive maps features, i.e., markers, overlapping marker spiderfier. Map tiles based on OpenStreetMap.

1.3.2 Plugins

In WordPress, a plugin is a PHP file that provides specific functionality to a website. It allows the theme to achieve a certain objective and help users tailor the website for their specific needs. Table 1-2 shows the list of plugins used for the Archive.

Table 1-2: List of plugins used in the Archive

Plugin	Description
Attributes Plugin	A plugin that is used to handle inappropriate content, ratings, and such. This feature was implemented into the system but is not being used.
Custom Email	Sends custom email from SHRP 2 Archive plugins, and adds a custom registration email
L13a Ingestion	Implements the custom file ingestion process for the L13a reliability data archive.
L13A WP-Admin Restriction Mode	Hides the WordPress Admin banner on top of the site.
Meteor Slides	Easily creates responsive slideshows with WordPress that are mobile friendly and simple to customize. In the SHRP 2 Archive system, the admin has the ability to insert a slideshow at the homepage.
S2 Comment Form	A plugin to add custom fields to the comment form.
SHRP 2 Custom Meta	This plug-in defines and enables custom metadata fields.
SHRP 2 Workflow	Enables administrators to manage artifacts in the SHRP 2 Archive.
Solr for WordPress	Provides access to the indexed content of the Solr search engine. Also, indexes existing Wordpress pages.
Theme My Login	Themes the WordPress login, registration and forgot password pages according to your theme.

1.4 MySQL Database

For most applications WordPress normally deals with the database by itself in such a way that the developer does not need to worry about the structure and the design of the database. The archive reads and writes data to tables in the MySQL database. Archive functionality is supported in MySQL by modifying existing WordPress tables and adding new tables. Section 7.3.1 and Section 7.3.2 of the L13A report review the native WordPress tables and the SHRP 2-specific tables in more detail.

Please note the Archive stores data sets in two formats:

- Data sets (CSV files) uploaded by users as part of the ingestion process are stored in the file system.
- In the ingestion process, the Archive stores the data sets in indexed database tables, which are used for visualizing and filtering data.

1.5 Solr Search Engine Server

Solr is an open source enterprise search engine project sponsored by the Apache Software Foundation. Solr provides keyword search functionality for the Archive. Solr is written in Java and runs as

a standalone full-text search server within a servlet container such as Tomcat. Solr uses the Apache Lucene Java search library at its core for full-text indexing and search, and has REST-like HTTP/XML and JavaScript Object Notation (JSON) APIs that make it easy to use from virtually any programming language (Apache Lucene 2014). The Archive's Solr engine has been installed on Apache Tomcat. Solr indexes any artifact and metadata being uploaded into the Archive before they become available on the Archive.

1.6 S2A SERVER

S2A server is a back-end server application written in Java that manages artifact workflow and processing states in the Archive. Depending on the type, an artifact goes through different back-end processes. The workflow controls various processing paths that an artifact goes through, from the time it is uploaded into the Archive till the moment it becomes available in (or gets deleted from) the Archive. S2A core functionalities are listed in Section 5.2 of the L13A report (Step 4. back-end processing).

There are three state variables by which the status of an artifact in the Archive is defined. These variables are stored in wp_posts table. Table 1-3 summarizes the state variables.

- **s2_wf_state** – Shows an artifact's workflow state. Figure 1-3 depicts the various workflow states.
- **s2_proc_state** – Indicates the back-end processing status of an artifacts. See Section 5.2 of the L13A report (Step 4. back-end processing) for more information.
- **s2_proc_msg** – Provides processing outcomes in a message for the Creator. The message is displayed on 'My Artifact' list located on 'My Profile' page.

Table 1-3: State Variables

State Variables	Description	Values
s2_wf_state	Workflow approval state	0=Ingest, the artifact is in the ingestion process but not yet submitted. This is the default state for a new artifact. 3=Unprocessed ● Triggered by: Submit button pressed in Step 4 of the ingestion process 2=Processing ● Triggered by: admin reviews and approves 1=Published, available for public use ● Triggered by: s2a server completes processing -1=Pre-trash ● Triggered by: admin action, sending to bin state -3=Trash ● Triggered by: s2a-server moves artifact from Gulag to Bin state -4=Processing error (validation, loading, or indexing) ● Triggered by: s2a server (see s2_proc_state and s2_proc_msg for details)
s2_proc_state	Processing state	-1=error

State Variables	Description	Values
		0=unprocessed (default) 1=validating 2=validation failed (see s2_proc_msg) 3=loading 4=load failed (see s2_proc_msg) 5=indexing 6=indexing failed (see s2_proc_msg) 10=complete
s2_proc_msg	Message for users from artifact processing	"Data set ingestion finished." "Could not parse XXX fields. Optimizing table." "Calculating column extents." "Internal error: unknown column type." "Failed to load."

The archive uses the concept of a Finite State Machine to manage the state of artifacts as they move through processing steps (Figure 1-3). The states for each artifact are stored in the database table 'wp_posts'. The columns 's2_wf_state' and 's2_proc_state' in this table store the workflow and processing states for each artifact, respectively. The workflow column (s2_wf_state) corresponds to the WorkflowState Java class in the S2A server. The processing state column (s2_proc_state) corresponds with the ProcessingState Java class in the S2A server. Figure 1-3 shows workflow states and the associated administrative (or processing) steps that cause the artifact to transition to a new state. The processing state is determined by the backend processing (S2A) of each artifact and may be one of the following: error, unprocessed, validating, validate_failed, loading, loading_failed, indexing, indexing failed, complete. Published artifacts have a processing state of complete (see the Java code for actual values). For administrative purposes, it may sometimes be useful to directly manipulate the states of artifacts via SQL update statements.

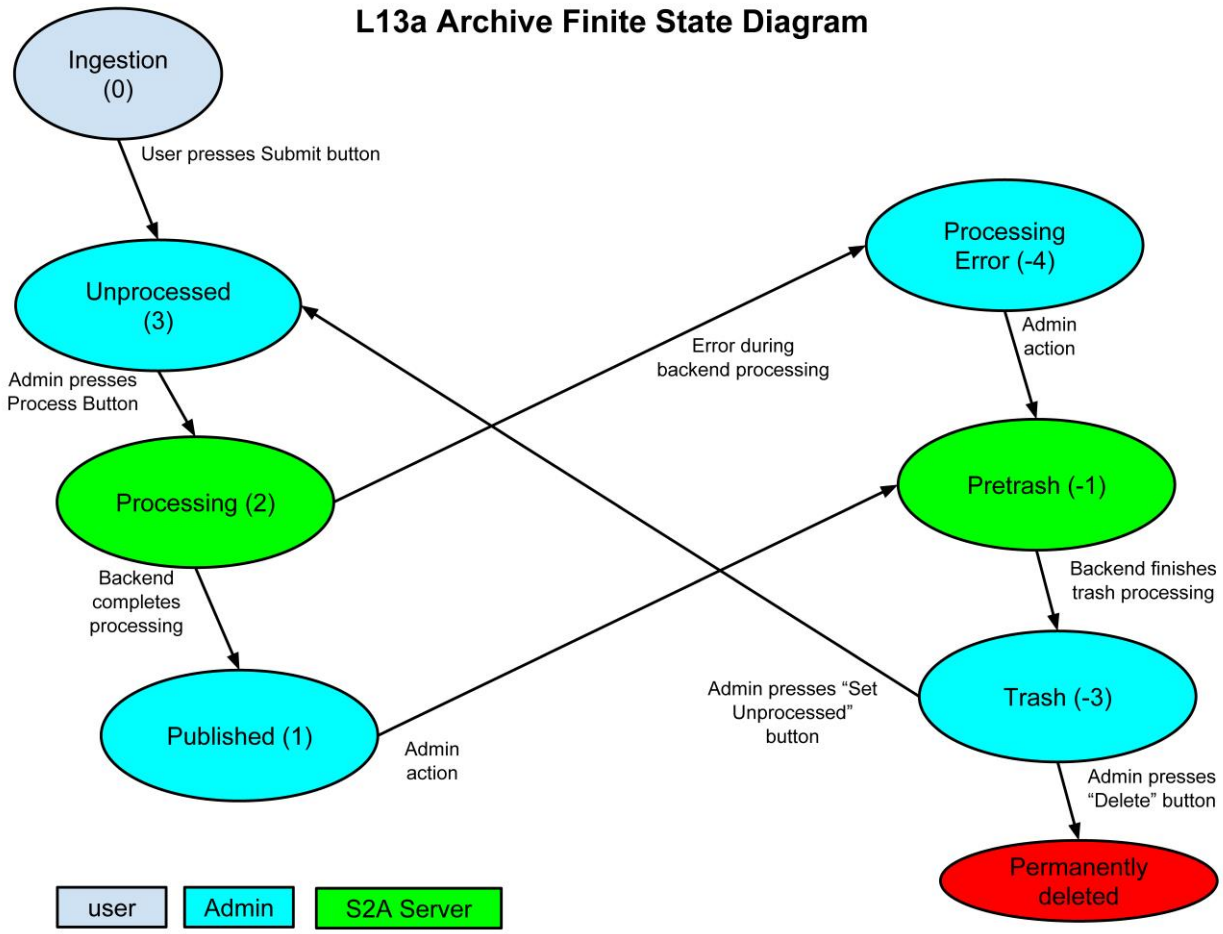

Figure 1-3: The Archive Processing Finite State Diagram

1.7 TRIPWIRE

Tripwire is an intrusion detection system that also provides integrity assurance, change management, and policy compliance. Tripwire can be used to detect unexpected changes to files. Tripwire scans take about 30 seconds on the L13A Archive System and generate reports showing changes to files.